Composite state machine miner

Composite State Machine Miner: Discovering and Exploring Multi-perspective Processes

Maikel L. van Eck*, Natalia Sidorova, and Wil M.P. van der Aalst

Eindhoven University of Technology, The Netherlands
{m.l.v.eck,n.sidorova,w.m.p.v.d.aalst}@tue.nl

Abstract. Process mining provides fact-based insights into processes based on behaviour captured in event data. An important aspect of this is the discovery of process models from such data. Traditionally, the focus of process discovery is on learning the ordering of activities. We deviate from this dominating activity view on processes to focus on states and state changes. Specifically, we aim to discover state-based models for processes where different facets, or perspectives, of the process can be identified. In this paper we describe an interactive process discovery tool that can be used to discover and explore state-based models for such multi-perspective processes: the *Composite State Machine Miner*. It quantifies and visualises the interactions between perspectives to provide additional process insights. This tool has been used to analyse the BPI Challenge 2012 data of a loan application process and product user behaviour data gathered by Philips during the development of a smart baby bottle equipped with various sensors.

Keywords: process discovery, state-based models, multi-perspective processes, interactive process exploration

1 Introduction

We assume that the reader is familiar with the basic concepts of process mining and process discovery, and we refer to [1] for an in-depth overview. The goal of most process discovery approaches is to obtain models that describe the ordering of activities within a process. These approaches usually have an implicit notion of the state of a process.

In this work we deviate from the dominating activity view on processes in order to focus explicitly on process states and state changes. This state view is more intuitive than an activity view for processes for which state information is explicitly available. Examples of such explicit state information are the diagnosis

* This research was performed in the context of the IMPULS collaboration project of Eindhoven University of Technology and Philips: "Mine your own body".

of a patient in a healthcare process or the status of an order in a purchasing process.

When studying the states of a process, that single process can have different facets, or *perspectives*, each with their own state space. For example, consider the homeostatic process in a person, parts of which regulate sleep and nutrition. From the perspective of sleep the state of a person can be e.g. awake or asleep, while the state of the nutrition perspective can be e.g. hungry, eating or sated. The state of a person at any point in time is the composition of the state of both perspectives. These perspectives have individual process cycles, but there are interdependencies between states from different perspectives, e.g. people are awake while eating. Our goal is to analyse these interdependencies between perspectives for multi-perspective state-based processes.

An approach to achieve this is described in detail in [3]. In this work we discuss the implementation of this approach: the *Composite State Machine (CSM) Miner*.

2 Implementation

The CSM Miner has been implemented as a plug-in of the ProM framework [4]. It is obtained by installing ProM 6.6 or later from `http://www.promtools.org/` and then using the ProM package manager to instal the CSMMiner package. Example logs that can be used with the CSM Miner can be obtained from `https://svn.win.tue.nl/repos/prom/Packages/CSMMiner/Logs`. There is a screencast providing a demonstration of the main features of the miner at: `https://svn.win.tue.nl/repos/prom/Packages/CSMMiner/Documentation`.

The main assumption behind the CSM Miner is that each state of the multi-perspective composite process under study is a vector of the states of its perspectives. That is, for a process with n perspectives each state of the composite process is a state of the form: $s = (s_1 \times \ldots \times s_n)$. The CSM Miner discovers a state machine describing the states and state changes of the composite process, as well as a state machine for each perspective. Transitions $(s \rightarrow s')$ in the discovered composite state machine represent state changes in one or more perspectives, while transitions $(s_i \rightarrow s'_i)$ in the discovered state machine of perspective i represent a change in that specific perspective.

The input data for the CSM Miner is assumed to be an XES event log [4] where each event represents a state change in a specific perspective. In Fig. 1 an example of the desired input is shown. Each trace in the log is required to contain attributes of the form `process:initialstate:[perspectiveName]` specifying the initial state of each perspective at the start of the trace. Each event is required to contain an attribute of the form `process:name` specifying the perspective for which it is a state change.

An example of a composite state machine and its perspective state machines discovered by the CSM Miner is shown in Fig. 3. Each state machine is displayed separately in an interactive visualisation where states and transitions can be dragged to move them around.

```
<log xes.version="1.0" xes.features="nested-attributes" openxes.version="1.0RC7">
    <string key="concept:name" value="Example Log"/>
    <trace>
        <string key="concept:name" value="Day1"/>
        <string key="process:initialstate:Feeding" value="Hungry"/>
        <string key="process:initialstate:Sleeping" value="Awake"/>
        <event>
            <string key="process:name" value="Feeding"/>
            <string key="concept:name" value="Eating"/>
            <string key="lifecycle:transition" value="start"/>
            <date key="time:timestamp" value="1970-01-01T06:00:00.000+01:00"/>
        </event>
        <event>
            <string key="process:name" value="Feeding"/>
            <string key="concept:name" value="Sated"/>
            <string key="lifecycle:transition" value="start"/>
            <date key="time:timestamp" value="1970-01-01T06:15:00.000+01:00"/>
        </event>
```

Fig. 1: A partial XES event log that can be used as input for the CSM Miner. Each trace in the log is required to be annotated with attributes specifying the initial state of each perspective and each event is required to be annotated with the perspective to which they belong.

States and transitions are annotated with statistics and additional statistics are shown at the bottom of the visualisation for the state or transition that is currently selected. For each state the statistics show the total number of times this state was observed to occur in the log, as well as the total time that was spent in this state. For each transition the statistics also show the total number of times this transition was observed. The percentage for each transition shows the observed frequency with which this transition was taken from the source state of the transition, i.e. if a state has only one outgoing transition then it is annotated with 100% and if there are two outgoing transitions that have been observed equally many times then they are each annotated with 50%.

By clicking on a state or transition the states and transitions in the other models that co-occur with the selected element are highlighted. The highlighted states also show two different statistics: confidence and lift. Confidence expresses what percentage of time the highlighted state was observed together with a selected state compared to the total time spent in that selected state. Lift expresses how much the confidence differs from the percentage of time that a highlighted state is expected to occur independent of the selected state. For a more detailed discussion of confidence and lift we refer to [3].

The visualisation of the discovered models also contains functionality that enables the user to transform the model. This functionality can be accessed from an expandable menu, as shown in Fig. 2, and allows the user to remove selected states from the model and to aggregate selected states. During the model transformation the statistics are recalculated.

3 Case Studies

The Composite State Machine Miner has been used during the analysis of two case studies. The first case study concerned the BPI Challenge 2012 data of

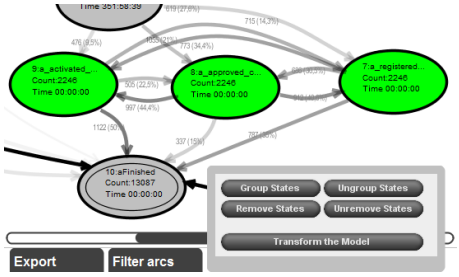

Fig. 2: The green states are selected for aggregation by the user because they are always executed together, but in arbitrary order.

a loan application process [2]. The second case study involved the analysis of product user behaviour data gathered by Philips during the development of a smart baby bottle equipped with various sensors. These case studies and the insights provided by the CSM Miner are discussed in detail in [3].

The models shown in Fig. 3 have been discovered on the BPI Challenge 2012 data. These models are much more structured than models discovered by traditional process discovery algorithms on the same data. Analysing them provided useful insights into the difference between automatically and manually processed applications and the effectiveness of fraud investigation.

4 Conclusion

In this paper we have presented the Composite State Machine Miner, an interactive tool for the discovery and exploration of state-based multi-perspective processes. The CSM Miner can be used to study the interactions between perspectives in such processes. This has been used successfully in two case studies.

Future work that we plan to do is focussed at improving the practical usability of the tool. For example, it would be useful to suggest states or relations between perspectives that are likely to be of interest to the user, to avoid having to click on every element of the discovered models to explore the results. Also, additional support for different types of relations between perspectives would enable the CSM Miner to generate additional insights.

References

1. van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer (2011)
2. van Dongen, B.F.: BPI Challenge 2012 (2012), http://dx.doi.org/10.4121/uuid:3926db30-f712-4394-aebc-75976070e91f
3. van Eck, M.L., Sidorova, N., van der Aalst, W.M.P.: Discovering and Exploring State-based Models for Multi-perspective Processes. In: 14th International Conference on Business Process Management (BPM). p. In press. Springer (2016)
4. Verbeek, H.M.W., Buijs, J.C.A.M., van Dongen, B.F., van der Aalst, W.M.P.: XES, XESame, and ProM 6. In: Information Systems Evolution, pp. 60–75 (2011)

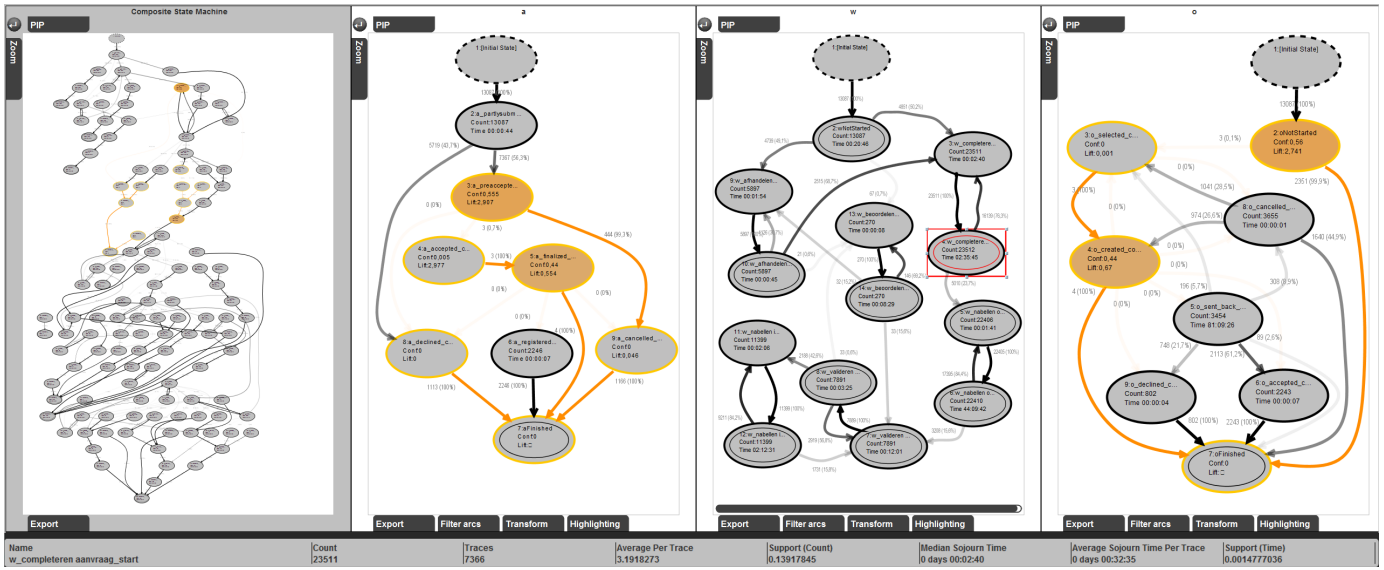

Fig. 3: The interactive visualisation of a discovered composite state machine. The selected state is denoted with a red box and its co-occurring states and transitions are highlighted in the other perspectives and the overall view based on their *confidence*. The statistics below the visualisation refer to the selected state or transition.